Dear Editor

I applaud the recent commentary by Dr Abers and Musher challenging the widespread use of the area under the receiver operator characteristic curve (AUC) to evaluate community-acquired pneumonia (CAP) severity indices. This statistic is very useful as a measure of discrimination for predicting an outcome (e.g. mortality) across a range of cut-offs, but says nothing about the clinical utility of the test. The clinical relevance depends critically on one question—what are we predicting and why?

While Abers and Musher rightly emphasise clinical utility over statistical validity, after 20 years of research into scores it would be nice to think we can achieve both. They suggest that sensitivity should be the key statistic in evaluating severity scores. Sensitivity is important, but I would like to suggest an alternative, the likelihood ratio, as potentially the most useful in clinical practice.

The test performance required depends most importantly on the population clinicians wish to identify. If identifying a low risk group of patients for discharge from the emergency department, clinicians should be primarily interested in the negative likelihood ratio—the degree to which the test will help you confidently exclude an adverse outcome such as mortality. Conversely, if deciding on intensive care unit (ICU) admission, the positive likelihood ratio will indicate the degree to which the test increases the pre-test probability of a poor outcome and therefore how much more likely it is that the patient will require ICU care.

Likelihood ratio is preferable to positive and negative predictive values that are too dependent on the prevalence of the outcome in the population.

Sensitivity alone cannot be used to guide clinical decisions. Taking the argument to its logical extreme, defining everyone with CURB65 score of 1 or more as ‘severe CAP’ achieves a sensitivity of 98.6% for mortality, therefore avoiding false negatives, but with a specificity of 26.5%, would require clinicians to treat nearly three quarters of hospitalized patients with CAP in the ICU.

Likelihood ratios, however, have the potential to achieve the goal correctly identified by Abers and Musher of incorporating pre-test probability and the score to arrive at a post-test probability that can be clinically interpreted. For example, if the patient has a pre-test probability of 10% for mortality from CAP, a CURB65 score of 4 gives a positive likelihood ratio of 5.4, and therefore a post-test probability of 38%. If the pre-test probability is 20%, the CURB65 score of 4 gives a post-test probability of mortality of 57%. This is potentially very valuable information for clinicians, much more so than the sensitivity or AUC.

I do not share the view that CURB65 should be abandoned. Physicians and guideline writers just need to have a sensible understanding of how to use these scores in clinical practice alongside the critical factors of pre-test probability and physician judgment.

J.D. Chalmers
Tayside Respiratory Research Group,
University of Dundee, Dundee, DD1 9SY, UK
e-mail: jchalmers@dundee.ac.uk

References

© The Author 2014. Published by Oxford University Press on behalf of the Association of Physicians. All rights reserved. For Permissions, please email: journals.permissions@oup.com
Response: Predicting poor outcomes in community-acquired pneumonia

We appreciate the gracious and insightful comments by Chalmers\(^1\) on our recent opinion paper.\(^2\) To fully consider the points raised by Chalmers, we decided to calculate the positive and negative likelihood ratio (+LR and –LR, respectively) for several scoring systems, using data from our recently published prospective study of community-acquired pneumonia (CAP) at the Michael E. DeBakey Veteran Affairs Medical Center. The details of this study are published elsewhere.\(^3\) Of 191 immunocompetent patients with CAP, 31 (16.2%) were admitted to the intensive care unit (ICU) and/or died within 30 days of admission. The performance of several CAP scores as predictors of a poor outcome (ICU admission and/or 30 day mortality) are shown in Table 1.

As Chalmers mentioned, –LR is of greater interest for clinicians, excluding low-risk patients (low pre-test probability of a poor outcome) from the emergency department, while +LR is better applied to high-risk (high pre-test probability of a poor outcome) patients when determining ICU admission. In our patients, SMART-COP and the IDSA/ATS minor criteria possessed the lowest –LRs and the highest +LRs. Similarly, a study by Chalmers et al.\(^4\) reported lower –LRs and higher +LRs for SMART-COP and the IDSA/ATS minor criteria compared with LRs for PSI and CURB-65.

In conclusion, we agree with Chalmers’ suggestion that LRs offer important advantages compared with other statistical parameters such as sensitivity and area under the receiver operator characteristic (AUROC). Furthermore, we cannot overemphasize Chalmers’ point about the importance of using clinical judgment as an adjunct to prediction rules.

Conflict of interest: None declared.

M.S. Abers
Department of Medicine, Massachusetts General Hospital, Boston, MA, USA
email: mabers@partners.org

N. Uy
Department of Medicine, Baylor College of Medicine, Houston, TX, USA

D.M. Musher
Department of Medicine, Baylor College of Medicine, Houston, TX, USA, Medical Care Line, Infectious Diseases Section, Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA

References

doi:10.1093/qjmed/hcu152

Advance Access Publication 30 July 2014

---

Table 1 Prediction of ICU admission and/or 30-day mortality\(^a\)

<table>
<thead>
<tr>
<th>Prediction rule</th>
<th>Sensitivity (95% CI)</th>
<th>+LR (95% CI)</th>
<th>–LR (95% CI)</th>
<th>AUROC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSI ≥ 4</td>
<td>71.0 (53.4–83.9)</td>
<td>1.18 (1.12–1.25)</td>
<td>0.73 (0.56–0.94)</td>
<td>0.65 (0.54–0.76)</td>
</tr>
<tr>
<td>CURB-65 ≥ 3</td>
<td>29.0 (16.1–46.6)</td>
<td>1.94 (1.05–3.58)</td>
<td>0.83 (0.76–0.92)</td>
<td>0.65 (0.55–0.75)</td>
</tr>
<tr>
<td>SMART COP ≥ 3</td>
<td>58.1 (40.8–73.6)</td>
<td>2.21 (1.95–2.51)</td>
<td>0.57 (0.49–0.67)</td>
<td>0.71 (0.60–0.82)</td>
</tr>
<tr>
<td>IDSA/ATS minor ≥ 3</td>
<td>51.6 (34.8–68.0)</td>
<td>3.44 (2.83–4.19)</td>
<td>0.57 (0.50–0.65)</td>
<td>0.71 (0.60–0.82)</td>
</tr>
</tbody>
</table>

\(^a\)Values reported with 95% confidence interval are given in parentheses.